SPEECH AS ARTICULATORY ENCODING OF COMMUNICATIVE FUNCTIONS

Yi Xu

Department of Phonetics and Linguistics, University College London, London
Haskins Laboratories, New Haven, USA
yi@phon.ucl.ac.uk

ABSTRACT

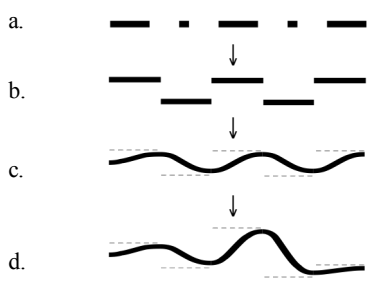
Speech conveys communicative meanings by encoding functional contrasts. The contrasts are realized through articulation, a biomechanical process with specific constraints. Phonology, phonetics or any other theories of speech therefore cannot be *autonomous* from either communicative functions or biophysical mechanisms. Successful speech modeling can be achieved only if communicative functions and biophysical mechanisms are treated as the *core* rather than the *margins* of speech.

Keywords: target approximation, parallel encoding, PENTA, unidirectionality

1. INTRODUCTION

Few would disagree that speech is unlike the Morse code, which encodes discrete symbolic information with a set of distinct long and short pulses, separated by distinct lengths of pauses. Speech is known, rather, to encode information in a much less discrete manner [13]. But it is an open question as to how different speech is from the Morse code. Imagine that we start with Morse-code-like short and long pulses (Fig. 1a) but replace them with tones of different frequencies, and remove all the pauses in between (Fig. 1b). Imagine further that, instead of a device that generates steady-state tones with abrupt onsets and offsets, we use a device whose state can be changed only sluggishly. What we will get is a continuous output like the solid curve in Fig. 1c, where the target tones are shown as the dashed lines.

Figure 1: From pseudo Morse code to continuous surface curves. See text for explanation.



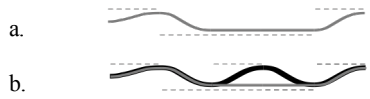
Imagine still further that we raise the third target and lower the fourth and fifth targets in Fig. 1c. The resulting output would then look like the curve in Fig. 1d. Now, if unaware of either the derivation history or the underlying components as indicated by the dashed lines, we would view the solid curve in Fig. 1d as neither discrete nor invariant.

In this paper I will argue that speech coding is actually not essentially different from the process in Fig. 1. I will demonstrate that such a coding process consists of two subprocesses: (a) target assignment and modification, and (b) target approximation, and that the entire process can be captured by a framework called *Parallel Encoding and Target Approximation* (PENTA) [36].

2. DIRECT CODING OR ARTICULATORY CODING?

If we were to understand the nature of the solid curve in Fig. 1c with no knowledge of its derivation history, a natural reaction would be to assume that the surface patterns are the code itself: consisting of rising and falling slopes or peaks and valleys, etc. We may also conclude upon further observation that the adjacent units overlap with each other, because no part of the signal seems to be exclusively influenced by any single unit. But without being told about the derivation history, how could we know otherwise? Suppose we know at least the identities of the coding elements and are able to manipulate them, say by making the third element identical to the two adjacent ones. We would then get the thin curve in Fig. 2a. Overlaying Fig. 2a with Fig. 1c we would get Fig 2b, from which we could see that, a) the difference in the middle part of Fig. 2b is only due to the third element, b) the third element has extensive influence on the portion of the curve corresponding to the fourth element, and c) but it has no influence on any of the preceding elements.

Figure 2: a. Same as Fig. 1c but with 3rd element identical to the surrounding elements. b. Overlay of a. and Fig. 1c.

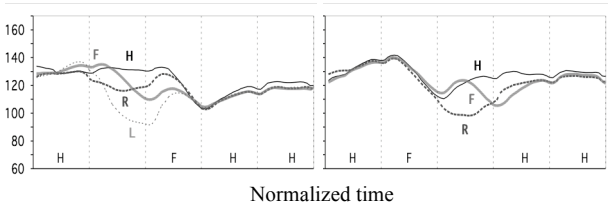


Interestingly, this is almost exactly what we have seen in lexical tones. Fig. 3 shows that in Mandarin, the tone of the second or third syllable in the 5-syllable utterances recorded in [34] has little influence on the preceding tone(s) but extensive influence on the following tone. Despite the influence, the F_0 curves of the 3rd syllable in

Fig. 3a gradually converge to a falling slope appropriate for the F tone. Likewise, the F_0 curves of the 4th syllable in Fig. 3b converge to a high-level shape appropriate for the H tone. Such convergence reveals a coding mechanism not unlike that seen in Fig. 1c. Several characteristics of the coding mechanism in Fig. 1-3 are worth noting:

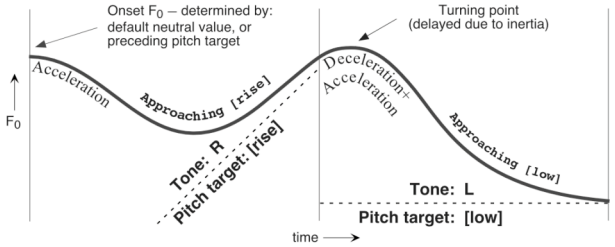
1. Unidirectionality — The surface curve is always moving monotonically toward one desired target or another.
2. No anticipatory execution — The movement toward a target does not start until the movement toward the preceding one is over.
3. No return to rest position — No portion of the curve is for the sake of returning to a non-target rest position after a target has been approached.

Figure 3: Mean F_0 contours of Mandarin five-syllable utterances. Adapted from [34].



This coding mechanism is captured by the Target Approximation model [41], as schematized in Fig. 4. Note that the greatest difference here from Fig. 1c is that the first target is a dynamic [rise], whose approximation results in a high velocity that forces the turning point to occur during the interval of the second target.

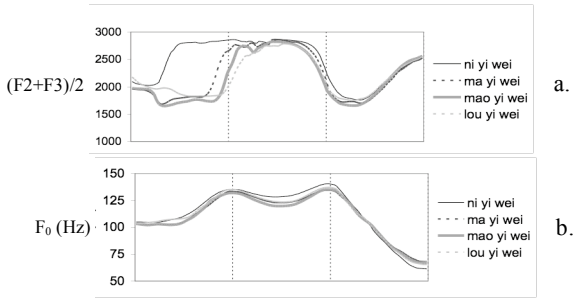
Figure 4: Illustration of the TA model. The vertical lines represent syllable boundaries. The dashed lines represent underlying pitch targets. The thick curve represents the F_0 contour that results from asymptotic approximation of the pitch targets.



There is evidence that sequential target approximation is involved also in the production of segmental units. Like F_0, what is critical is to manipulate the segmental context while keeping the target segment constant. In Fig. 5a each curve is an average of F2 and F3 of trisyllabic Mandarin phrases. The second syllable is [ji] while the first syllable is [ni], [ma], [mao] or [lou]. They are time-aligned according to their F_0 contours (Fig. 5b) which all have the tone pattern of RRL. Such time alignment is justified by the finding that certain F_0 events are consistently aligned to segmental events [16, 34, 42]. The differences due to the first syllable seems to go only as far as before the mid point between the two F_0 peaks. This indicates that the implementation of the glide [j] is a process of target approximation similar to that of tones described above.

In Fig. 6, the same syllable is followed by either [weɪ] or [joʊ]. They are also time-aligned to F_0 peaks in the same manner as Fig. 5. Here the point of formant divergence due to the third syllable occurs at about 1/4 or 1/5 of the middle interval (as marked by the F_0 peaks), which is on average 50 ms before the second F_0 peak. As found in [34, 35], the F_0 peak of a R tone occurs about 28-29 ms after the nasal murmur onset of the following syllable with a L tone. This means that the formant movements toward the initial [j] or [w] start no earlier than 50 ms before the equivalent of the nasal murmur onset [38].

Figure 5: Carryover effect on formants. a. LPC-tracked mean formant curves [(F2+F3)/2] of trisyllabic Mandarin phrases, averaged across 10 repetitions by a male speaker. Time normalization is done in the three intervals divided by the two F_0 peaks shown in b. b. Mean F_0 curves of the same phrases.



The patterns of contextual formant variations in Fig. 5-6 are in fact remarkably similar to a series of findings reported in the 80s and early 90s, as summarized in [4], which concludes that "there is actually rather little 'anticipation' of articulatory activity." [4] concludes further that "segmental articulatory behaviors do not extend very far from the segments for which they were intended." But note that here segments are still defined in terms of acoustic landmarks.

What if the segmental interval is also defined in terms of target approximation? Then the F2-3 movements toward [w] and [j] in syllable 3 in Fig. 6 are no longer their *anticipation*, but their *execution*. In other words, the acoustic onset of a segment is actually much earlier than its landmark-defined conventional onset. By the same token, a vowel interval could be defined as starting from the beginning of the articulatory/acoustic movement toward its target. If so, in a CV syllable, V actually starts at about the same time as C, based on findings reported as early as 1933 and 1966 [15, 18, 20]. This can be seen in Fig. 7, where the F2 movements toward the vowel targets of the second syllable start not after the offset of the [l] murmur, but well before the onset of the [l] murmur, going downward toward [u] in Fig. 7a, but upward toward [i] in Fig. 7b. Note that because the movements start when

the spectral patterns are typical of the first vowel, they are conventionally viewed as due to coarticulation [8]. Under the definition of unidirectional target approximation, however, they are viewed as part of the second vowel.

Figure 6: Anticipatory effect on formants. Mean formant curves similar to those of Fig. 5.

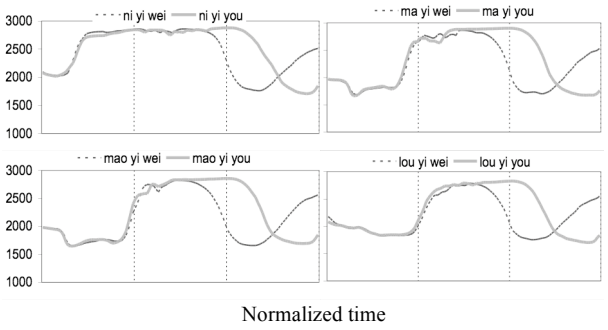
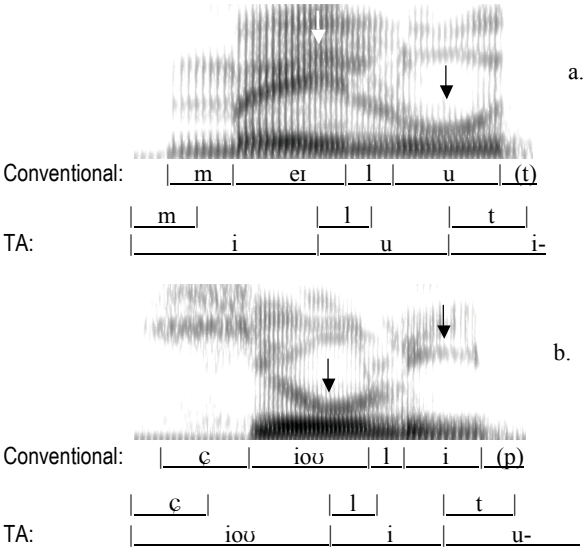


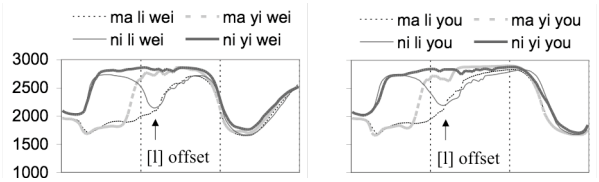
Figure 7: Spectrograms and conventional vs. TA segmentations of Mandarin [l] + V sequences. a. [meɪ lu (tiɛn χuo)] (to light coal stove). b. [ɕiou li (bu tʂou)] (repair procedure). The arrows mark F2 turning points.



Despite the CV co-onset, the movements of any particular articulator shared by both C and V are still sequential [33]. In Fig. 8, for example, F2-3 of /ma li/ and /ni li/ finishes approaching the [l] target before starting to approach the [i] target. Note that such sequential nature of articulation is hard to reveal by isolated examples, as those shown in [28]. It is essential to make observations based on minimal pairs [cf. 4].

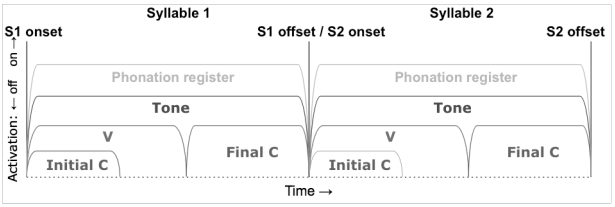
The above understanding has led to the time structure model of the syllable [39], according to which the syllable serves as a time structure that assigns the temporal intervals of consonants, vowels, tones and phonation registers, as illustrated in Fig. 9. The alignment is hypothesized to follow three principles: a) Co-onset of the initial consonant, the first vowel, the tone and the phonation register at the beginning of the syllable, b) Sequential offset of all non-initial segments, especially coda C, and c) Synchrony of tone and phonation register with the entire syllable.

Figure 8: Sequential C-V articulation. LPC-tracked mean formant curves [(F2+F3)/2] of Mandarin phrases, averaged across 10 repetitions by a male speaker.



The discussion so far thus points to an emerging picture that speech coding is much more sequential than conventionally thought, at least at the most basic level. The picture also allows us to see that, without knowing the articulatory mechanisms, it is difficult, if not impossible, to determine from surface forms what the underlying components are like. As I will show next, however, an even clearer picture can be seen only when we treat communicative functions also as part of the core rather than the margins of speech.

Figure 9: The time structure model of the syllable. Adapted from [39].



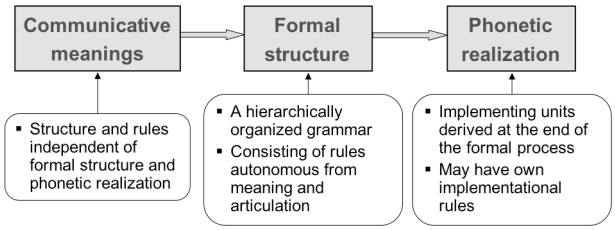
3. FORM FOR FORM'S SAKE OR FORM AS CODING FOR FUNCTIONS?

For a long time, the dominant view about speech production is that it consists of three separate processes, as schematized in Fig. 10. First, the communicative meanings form their own structure and rules independent of other processes (left). They are then passed on to a formal process that is autonomous from both meanings and articulation (middle). Autonomous because it implements a grammar whose function is to define what is *well-formed* and what is not [6]. It is after the completion of this formal process that the output of the formal derivation is passed on to the phonetic process (right), which also has its own rules [3, 23, 24]. Thus the task of speech production is first and foremost to guarantee that a set of formal requirements are met. And understanding speech is done not by directly accessing the meanings, but by first parsing a formal structure [6].

There are at least two fundamental problems with this view, due to its neglect of two basic facts: that speech is articulatorily generated, and that it conveys communicative meanings. First, by assuming autonomy from articula-

tion, the link between underlying units and surface form is explained in terms of inductively derived stipulations or markedness constraints [9]. This is apparently circular, for it tries to explain the observed variations directly in terms of observed tendencies themselves [19]. Secondly, by assuming autonomy from meaning, phonological units have to be postulated based solely on surface forms. The approach is thus by nature error-prone, for it lacks external references. As has been demonstrated previously, ignoring the meaning-form link even temporarily has led to a failure to recognize the true functional units [42, 36, 37]. In particular, pitch accents, widely accepted as the basic units in intonation, have an inter-labeler consistency of no more than 50% even by experts performing repeated visual and auditory examinations [32].

Figure 10: A schematic of conventional conceptualization of the speech production process.



Both problems in fact have been pointed out before, and various alternatives have been suggested [e.g., 12, 14, 19, 26]. But what has been lacking is an alternative that can address both at the same time, by establishing a continuous link between communicative functions and surface acoustics through articulation. This has motivated the PENTA model [36], originally proposed for tone and intonation, but here extended to other aspects of speech. A diagram of PENTA is shown in Fig. 11. The stacked boxes on the far left represent individual communicative functions as the primary input to the model. They are parallel to each other with no hierarchical organizations. The communicative functions are realized through distinctive encoding schemes (second stack of boxes from left), which are either universal or language specific. The encoding schemes specify the values of the target approximation parameters (middle block). These parameters then control the articulatory process of syllable-aligned sequential target approximation (right) to generate surface acoustic output. In the following, I will discuss the functional and mechanistic aspects of PENTA separately.

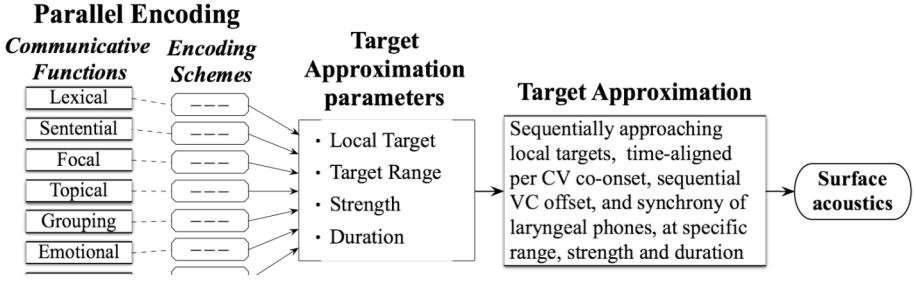
3.1. Communicative functions from a broader perspective

It is widely held that there is a critical division between linguistic and paralinguistic aspects of speech, although there is little agreement on the exact location of the dividing line. There have been proposals to expand the linguistic domain to include non-lexical functions [23, 24], but the insistence on the division continues [cf. 3]. From a functional perspective, this assumed division is not very informative, as any meaning being deliberately conveyed is functional. Speakers convey to the listeners not only words, but also many layers of non-lexical and non-syntactic information. Instead of assuming a single linguistic/paralinguistic division, it is much more important to recognize specific divisions between individual communicative functions. Although not all the functions are as apparent as the lexical function, their recognition may be made easier by following the principles proposed in [37]:

1. *Specificity*. Communicative functions should be as specific as possible about what they contrast and about what their temporal domains of operation are.
2. *Mutual-exclusivity*. Each function should have a unique "*encoding scheme*" which has at least one predominant characteristic not overlapped by other functions.
3. *Audibility*. A functional contrast in a language must have reached certain perceptual threshold, otherwise it would not have been operational.
4. *Elicitability*. For a function to be verifiable, there needs to be at least one way of reliably eliciting it under experimental conditions.

Also from a functional perspective, many conventional issues may be viewed in a very different light. First, not all communicative functions have to be categorical. Instead, the degree of categoricalness should depend on the nature of the function. Lexical functions, including lexical tones, are necessarily categorical because word identity is by nature unequivocal. In contrast, pitch reset, which likely serves to introduce a new topic or a new turn in a conversation, has been found to be quite gradient, with close correlation with perceptual judgment of the size of the break

Figure 11: A schematic sketch of the general PENTA model. Modified from [36].



between two utterances [30]. While the exact nature of the function is still not fully clear, it is possible that there are no fixed categories of breaks as assumed in the ToBI convention [40]. As long as its meaning and encoding scheme are clearly definable, a gradient function can still operate in parallel with other, more categorical functions.

Second, whether the form-meaning association is arbitrary or iconic no longer needs to be viewed as a determining factor for the nature of a function. As is well known, the iconicity of onomatopoeia is only superficial, for different languages may use onomatopoeic words for the same phenomenon, but in very different iconic ways. The same should be true for non-lexical functions. The seemingly iconic encoding scheme for focus found in Mandarin, English and many other languages, as will be further discussed later, may appear perfectly natural. But there is emerging evidence that there are languages that do not encode focus in quite the same way [21]. Likewise, the also seemingly natural fall/rise contrast between statement and question may be totally missing in a group of African languages that employ means such as final lengthening and breathiness to indicate questions [27].

On the other hand, the degree of iconicity itself may not be totally arbitrary. Vowels, consonants and tones are mostly non-iconic in lexical encoding because, with only a few exceptions, it is not easy to associate meanings to vocal articulatory states. In contrast, iconicity is more common in sign languages because meaning-gesture links can be much more easily made [25]. But signs, despite their iconicity, are just as arbitrary as onomatopoeia [25]. Also, the extent to which biological code [10] forms the basis of a particular encoding scheme depends much on the internal crowdedness of the function: The focal function, which can be viewed as quite iconic, contrasts only between focused and non-focused items. The lexical function, however, contrasts tens of thousands of words from each other with only a few dozen vocal tract shapes and their dynamic patterns. Thus the nature of the function and its available encoding space probably jointly determine its degree of iconicity.

Third, communicative functions are by nature orthogonal to each other, and so there is unlikely a cross-functional hierarchy that governs all functions, as often assumed [7, 24]. In English, for example, word stress serves a lexical function. Focus, on the other hand, is to highlight a particular element against other elements in an utterance, and so is assigned independently of lexical stress. Phonetically, focus neither creates nor eliminates F_0 peaks and valleys that already exist in a neutral-focus utterance, but instead only modifies their pitch ranges [42]. Likewise, syllable grouping is probably encoded in Mandarin by duration adjustments without the mediation of stress [40]. Thus the grouping function is again orthogonal to both lexical stress and focus.

Finally, the manner with which multiple functions are encoded means that phonetic variability is inevitable if the coding for non-lexical functions is treated as noise [22]. If, however, as many involved functions as can be recognized are controlled, random variability may be actually much smaller than has often been reported.

3.2. Encoding communicative functions via Target Approximation

Given that so many communicative functions need to be encoded in speech, how is it possible to implement them in a manner that guarantees perceptual decoding? The solution postulated by the PENTA model is that the process of *syllable-aligned target approximation* serves as the *base mechanism* of speech coding, through which all layers of information are encoded by manipulating the TA parameters, including pitch target, pitch range, strength and duration, as detailed in [36].

Variations due to the target assignment are apparent. What needs to be added is that the assignment of a target for each and every articulator at any particular interval is obligatory. That is, contrary to what is argued from a perceptual perspective [2], there cannot be articulatory underspecification anywhere, as the only assumed mechanism of controlling the state of an articulator is to make it approach a specific target. Note that articulatory specifications do not necessarily always mean functional specifications. In fact, the encoding scheme of a function typically involves specifications for only a limited number of articulators. When there is no functional specification for any of the parameters, it is likely that a default neutral value is assigned [5].

In addition to the target, parameters that determine its manner of approximation can also be functionally specified. Focus, for example, involves a tri-zone pitch range manipulation in English and Mandarin: expansion of the focused element, compression of the post-focus elements, and neutral on the pre-focus elements [34, 42]. Pitch range specification has also been found to be part of the encoding scheme of sentence type [17], and new topic [31].

Articulatory strength determines the speed at which a target is approached. It has been found to be used as part of the encoding scheme of the neutral tone in Mandarin [5]. This, among other things, has been interpreted as indication that the weak articulatory strength is lexically assigned to the neutral tone to contrast it with other tones [5]. Similar weak strength has been found in weak lexical stress in English [42].

Finally, target duration determines how much time is allocated to the approximation of a target. Durational differences is known to be used in many languages to contrast words [1, 11, 29]. The neutral tone syllables are about 61% as long as other syllables in Mandarin [5]. Durational differences have also been found to encode syllable grouping [40], as mentioned earlier.

4. CONCLUDING REMARKS

Conventional theories try to explain speech by assuming that speakers are guided by a set of symbolic rules — a grammar [2, 3, 6]. More recent theories assume that speakers are guided by rules which themselves are derived from perceptual and articulatory constraints [9]. None of these theories would deny, however, that speech also conveys meanings. But none of them has seriously contemplated the possibility that conveying meanings is all speech is about. In this paper I have explored this possibility by presenting the articulatory-functional view of speech, as embodied in the PENTA model [36]. According to this view, communicative functions are the actual driving force of speech, and their link to the acoustic signals is through encoding schemes that directly control articulation rather than through a self-contained phonological structure. The consequence of applying this view is that much of the observed variability can be traced to either functional or articulatory sources, and thus no longer needs to be treated as noise.

5. REFERENCES

[1] Abramson, A. S., Ren, N. 1990. Distinctive vowel length: Duration versus spectrum in Thai. *J. Phon*. 18, 79-92.

[2] Aditi, L. 2007. Non-equivalence between phonology and phonetics. This session.

[3] Arvaniti, A. 2007. On the relationship between phonology and phonetics (or why phonetics is not phonology). This session.

[4] Bell-Berti, F., Krakow, R. A., Gelfer, C. E., Boyce, S. 1995. Anticipatory and carryover effects: Implications for models of speech production. In *Producing Speech: Contemporary Issues. For Katherine Safford Harris*. F. Bell-Berti and L. J. Raphael. (eds.) New York: AIP Press.

[5] Chen, Y., Xu, Y. 2006. Production of weak elements in speech -- Evidence from f0 patterns of neutral tone in standard Chinese. *Phonetica* 63, 47-75.

[6] Chomsky, N., Halle, M. 1968. *The Sound Pattern of English*. New York: Harper & Row.

[7] Goldsmith, J. A. 1990. *Autosegmental and Metrical Phonology*. Oxford: Blackwell Publishers.

[8] Hardcastle, W. J., Hewlett, N., Eds. (1999). *Coarticulation: Theory, Data and Techniques*. Cambridge, Cambridge University Press.

[9] Hayes, B., Kirchner, R., Steriade, D. 2004. *Phonetically Based Phonology*. Cambridge: Cambridge Univ. Press.

[10] Hinton, L., Nichols, J., Ohala, J. J. 1995. *Sound symbolism*. Cambridge: Cambridge University Press.

[11] Hirata, Y. 2004. Effects of speaking rate on the vowel length distinction in Japanese. *J. Phon*. 32, 565-589.

[12] Hirst, D. J. 2005. Form and function in the representation of speech prosody. *Speech Commun*. 46, 334-347.

[13] Hockett, C. F. 1955. *A manual of phonology (International J. Am. Linguistics, Memoir 11)*. Baltimore: Waverly Press.

[14] Kohler, K. J. 2004. Prosody Revisited — FUNCTION, TIME, and the LISTENER in Intonational Phonology. *Proc. Speech Prosody 2004*, Nara, Japan, 171-174.

[15] Kozhevnikov, V. A., Chistovich, L. A. 1965. *Speech: Articulation and Perception*. Washington, DC: Joint Publications Research Service.

[16] Ladd, D. R., Faulkner, D., Faulkner, H., Schepman, A. 1999. Constant "segmental anchoring" of F0 movements under changes in speech rate. *J. Acoust. Soc. Am*. 106, 1543-1554.

[17] Liu, F., Xu, Y. 2005. Parallel Encoding of Focus and Interrogative Meaning in Mandarin Intonation. *Phonetica* 62, 70-87.

[18] Menezes, C. Menzerath, P., de Lacerda, A. 1933. *Koartikulation, Seuerung und Lautabgrenzung*. Berlin and Bonn: Fred. Dummlers.

[19] Ohala, J. J. 1990. There is no interface between phonology and phonetics: a personal view. *J. Phon*. 18, 153-171.

[20] Öhman, S. E. G. 1966. Coarticulation in VCV utterances: Spectrographic measurements. *J. Acoust. Soc. Am*. 39, 151-168.

[21] Pan, H. 2007. Focus and Taiwanese unchecked tones. In *Topic and Focus: Cross-linguistic Perspectives on Meaning and Intonation*. C. Lee, M. Gordon and D. Büring. Springer**:** 195-213.

[22] Perkell, J. S., Klatt, D. H., Eds. (1986). *Invariance and variability of speech processes*. Hillsdale, NJ, LEA.

[23] Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation. MIT, Cambridge, MA.

[24] Pierrehumbert, J. 1990. Phonological and Phonetic representation. *J. Phon*. 18, 375-394.

[25] Pietrandrea, P. 2002. Iconicity and Arbitrariness in Italian Sign Language. *Sign Language Studies* 2, 296-321.

[26] Port, R. F. 2005. Against Formal Phonology. *Language* 81, 927-964.

[27] Rialland, A. in press. African "lax" question prosody: its realisations and its geographical distribution. To appear in *Lingua*.

[28] Shih, C., Kochanski, G. 2007. Partial reduction, look-ahead and orderly variability in speech. This session.

[29] Suomi, K. 2005. Temporal conspiracies for a tonal end: Segmental durations and accentual f0 movement in a quantity language. *J. Phon*. 33, 291-309.

[30] Swerts, M. 1997. Prosodic features at discourse boundaries of different length. *J. Acoust. Soc. Am*. 101, 514-521.

[31] Wang, B., Xu, Y. 2006. *Prosodic encoding of topic and focus in Mandarin*. In *Proceedings of Speech Prosody 2006*, Dresden, Germany. pp. PS3-12_0172.

[32] Wightman, C. W. 2002. ToBI or not ToBI. *Proc. Speech Prosody 2002*, Aix-en-Provence, France. pp. 25-29.

[33] Wood, S. A. J. 1996. Assimilation or coarticulation? Evidence from the temporal co-ordination of tongue gestures for the palatalization of Bulgarian alveolar stops. *J. Phon*. 24, 139-164.

[34] Xu, Y. 1999. Effects of tone and focus on the formation and alignment of f_0 contours. *J. Phon*. 27, 55-105.

[35] Xu, Y. 2001. Fundamental frequency peak delay in Mandarin. *Phonetica* 58, 26-52.

[36] Xu, Y. 2005. Speech Melody as articulatorily implemented communicative functions. *Speech Commun*. 46, 220-251.

[37] Xu, Y. 2006. *Speech prosody as articulated communicative functions*. In *Proceedings of Speech Prosody 2006*, Dresden, Germany. pp. SPS5-4-218.

[38] Xu, Y., Liu, F. 2007. Determining the temporal interval of segments with the help of F_0 contours. *J. Phon*. 35, 398-420.

[39] Xu, Y., Liu, F. in press. Tonal alignment, syllable structure and coarticulation: Toward an integrated model. To appear in *Italian J. Linguistics*.

[40] Xu, Y., Wang, M. 2005. Tonal and durational variations as phonetic coding for syllable grouping. *J. Acoust. Soc. Am*. 117, 2573.

[41] Xu, Y., Wang, Q. E. 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Commun*. 33, 319-337.

[42] Xu, Y., Xu, C. X. 2005. Phonetic realization of focus in English declarative intonation. *J. Phon*. 33, 159-197.